```
1 CGGGGCCGAG GGCGGCGTCG CTGAGGCGCC CATGGCCTTC GCCCGCCGGC
   51 TCCTGCGCGG GCCACTGTCG GGGCCGCTGC TCGGGCGGCG CGGGGTCTGC
  101 GCTGGGGCCA TGGCTCCGCC GTGCCGCTTC GTCCTGGAGC TTCCCGACTG
  151 CACCCTGGCT CACTTCGCCC TAGGCGCCGA CGCCCCCGGC GACGCAGACG
  201 CCCCCGACCC CCGCCTGGCG GCGCTGTTGG GGCCCCCGGA GCGCAGCTAC
  251 TCGCTGTGCG TGCCCGTGAC CCCGGACGCC GGCTGCGGGG CCCGGGTCCG
  301 GGCGGCGCGG CTGCACCAGC GCCTGCTGCA CCAGCTGCGC CGCGGCCCCT
  351 TCCAGCGGTG CCAGCTGCTC AGGCTGCTCT GCTACTGCCC GGGCGGCCAG
  401 GCCGGCGGCG CACAGCAAGG CTTCCTGCTG CGCGACCCCC TGGATGACCC
  451 TGACACCCGG CAAGCGCTGC TCGAGCTGCT GGGCGCCTGC CAGGAGGCAC
  501 CACGCCCGCA CTTGGGCGAG TTCGAGGCCG ACCCGCGCGG CCAGCTGTGG
  551 CAGCGCCTCT GGGAGGTGCA AGACGGCAGG CGGCTGCAGG TGGGCTGCGC
  601 ACAGGTCGTG CCCGTCCCGG AGCCCCCGCT GCACCCGGTG GTGCCAGACT
  651 TGCCCAGTTC CGTGGTCTTC CCGGACCGGG AAGCCGCCCG GGCCGTTTTG
  701 GAGGAGTGTA CCTCCTTTAT TCCTGAAGCC CGGGCAGTGC TTGACCTGGT
  751 CGACCAGTGC CCAAAACAGA TCCAGAAAGG AAAGTTCCAG GTTGTTGCCA
  801 TCGAAGGACT GGATGCCACG GGTAAAACCA CGGTGACCCA GTCAGTGGCA
  851 GATTCACTTA AGGCTGTCCT CTTAAAGTCA CCACCCTCTT GCATTGGCCA
  901 GTGGAGGAAG ATCTTTGATG ATGAACCAAC TATCATTAGA AGAGCTTTTT
  951 ACTCTTTGGG CAATTATATT GTGGCCTCCG AAATAGCTAA AGAATCTGCC
 1001 AAATCTCCTG TGATTGTAGA CAGGTACTGG CACAGCACGG CCACCTATGC
 1051 CATAGCCACT GAGGTGAGTG GGGGTCTCCA GCACCTGCCC CCAGCCCATC
 1101 ACCCTGTGTA CCAGTGGCCA GAGGACCTGC TCAAACCTGA CCTTATCCTG
 1151 CTGCTCACTG TGAGTCCTGA GGAGAGGTTG CAGAGGCTGC AGGGCCGGGG
 1201 CATGGAGAAG ACCAGGGAAG AAGCAGAACT TGAGGCCAAC AGTGTGTTTC
 1251 GTCAAAAGGT AGAAATGTCC TACCAGCGGA TGGAGAATCC TGGCTGCCAT
 1301 GTGGTTGATG CCAGCCCCTC CAGAGAAAAG GTCCTGCAGA CAGTATTAAG
 1351 CCTAATCCAG AATAGTTTTA GTGAACCGTA GTTACTCTGG CCAGGTGCCA
 1401 CGTCTAACTA GATTAGATGT TGTTTGAAAC ATCTACATCC ACCATTTGTT
 1451 ATGCAGTGTT CCCAAATTTC TGTTCTACAA GCATGTTGTG TGGCAGAAAA
 1501 CTGGAGACCA GGCATCTTAA TTTTACTTCA GCCATCGTAC CCTCTTCTGA
 1551 CTGATGGACC CGTCATCACA AAGGTCCCTC TCATCATGTT CCAGTGAGAG
 1601 GCCAGCGATT GCTTTCTTCC TGGCATAGTA AACATTTTCT TGGAACATAT
 1651 GTTTCACTTA ATCACTACCA AATATCTGGA AGACCTGTCT TACTCAGACA
 1701 GCACCAGGTG TACAGAAGCA GCAGACAAGA TCTTCCAGAT CAGCAGGGAG
 1751 ACCCCGGAGC CTCTGCTTCT CCTACACTGG CATGCTGATG AGATCGTGAC
 1801 ATGCCCACAT TGGCTTCTTC CACATCTGGT TGCACTCGTC ATGATGGGCT
 1851 CGCTGCATCT CCCTCAGTCC CAAATTCTAG TAGCCAAGTG TTCCTGCAGA
 1901 GGCTGTCTAT GTGTCCTGGC TGCCCAAGGG ACACTCCTGC AGAGCCATTT
 1951 TTGGGTAAGG AACACTTACA AAGAAGGCAT TGATCTTGTG TCTGAGGCTC
 2001 AGAGCCCTTT TGATAGGCTT CTGATGTCAT TCATAAAGAC ATTCAAGCCA
 2051 AGATGCTCCA ACTGCAAATA TACCAACCTT CTCTGAATTA TATTTTGCTT
 2101 ATTTATATTT CTTTTCTTTT TTTCTAAAGA ATTGGCTCTG AATAGAATGC
 2151 ACATTTTCCA TCTGAACTGG ATGCATATCA TTTAGCCAAT CCAGTAATTT
 2201 ATTTATATTA ATCTATACAT AATATGTTTC CTCAGCATAG GAGCTATGAT
 2251 TCATTAATTA AAAGTGGAGT CAAAACGCTA AATGCAATGT TTGTTGTGTA
 2301 TTTTCATTAC ACAAACTTAA TTTGTCTTGT TAAATAAGTT CAAGTGGATC
 2351 TTGGAGTGGG ATTTCTTGGT AAATTATCTT GCACTTGAAT GTCTCATGAT
 2401 TACATATGAA ATCGCTTTGA CATATCTTTA GACAGAAAAA AGTAGCTGAG
 2451 TGAGGGGGAA ATTATAGAGC TTGTGTGACT TTAGGGAGTA GCTGTCTCTT
 2501 ATACACATAC TCAAGCCCTG AAGCCTTGCA TGTCCTGCAG CGTCGCACTA
 2551 AAGGAGGGGG CTTTTGCACC C
      (SEQ ID NO: 1)
```

FEATURES:

```
5'UTR:       1 - 31
Start Codon: 32
Stop Codon:  1379
3'UTR:       1382
```

Homologous proteins:

FIG.1A

Sequences producing significant alignments:	Score (bits)	E Value
CRA|98000043611838 /altid=gi|12835871 /def=dbj|BAB23396.1| (AK0...	608	e-173
CRA|18000004977190 /altid=gi|10181118 /def=ref|NP_065582.1| thy...	262	1e-68
CRA|335001098644584 /altid=gi|11466103 /def=ref|NP_047035.1| TK...	116	1e-24
CRA|67000041026087 /altid=gi|13562008 /def=gb|AAK30607.1|AF3502...	81	6e-14
CRA|18000004875826 /altid=gi|115398 /def=sp|P17656|CC02_CAEEL C...	80	1e-13
CRA|67000041026083 /altid=gi|13562004 /def=gb|AAK30605.1|AF3502...	80	2e-13
CRA|18000004994956 /altid=gi|71405 /def=pir||CGCH1S collagen al...	78	4e-13
CRA|18000004942943 /altid=gi|115268 /def=sp|P02457|CA11_CHICK C...	78	4e-13
CRA|271274599 /altid=gi|7801278 /def=emb|CAB91142.1| (AL355913)...	78	4e-13
CRA|18000005169791 /altid=gi|3641659 /def=dbj|BAA33381.1| (AB00...	77	9e-13

dbEST:

CRA Number	gi Number	Score	Expect
CRA|157000141080761	gi|13462554	1360 bits (686)	0.0
CRA|78000169053550	gi|14048666	1273 bits (642)	0.0
CRA|78000169051306	gi|14048462	1221 bits (616)	0.0
CRA|155000041537799	gi|10162440	1193 bits (602)	0.0
CRA|156000152698673	gi|11155283	1154 bits (582)	0.0
CRA|156000152689550	gi|11154451	1146 bits (578)	0.0
CRA|78000169034399	gi|14046924	1118 bits (564)	0.0
CRA|224000004311810	gi|15928374	1102 bits (556)	0.0
CRA|157000141097724	gi|13462915	1063 bits (536)	0.0
CRA|78000169034718	gi|14046953	1059 bits (534)	0.0
CRA|155000041535193	gi|10162203	1045 bits (527)	0.0
CRA|58000098794394	gi|12765577	969 bits (489)	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

gi Number	Organ	Tissue Type
gi|13462554	prostate	adenocarcinoma
gi|14048666	prostate	adenocarcinoma
gi|14048462	prostate	adenocarcinoma
gi|10162440	bone marrow	from acute myelogenous leukemia
gi|11155283	bone marrow	from acute myelogenous leukemia
gi|11154451	bone marrow	from acute myelogenous leukemia
gi|14046924	lung	mucoepidermoid carcinoma
gi|15928374	pooled brain, lung, testis	(none)
gi|13462915	prostate	adenocarcinoma
gi|14046953	lung	mucoepidermoid carcinoma
gi|10162203	bone marrow	from acute myelogenous leukemia
gi|12765577	prostate	adenocarcinoma, cell line

Tissue expression:
Pooled: Brain, Heart, Kidney, Lung, Spleen, Testis, Leukocyte

FIG.1B

```
   1 MAFARRLLRG PLSGPLLGRR GVCAGAMAPP CRFVLELPDC TLAHFALGAD
  51 APGDADAPDP RLAALLGPPE RSYSLCVPVT PDAGCGARVR AARLHQRLLH
 101 QLRRGPFQRC QLLRLLCYCP GGQAGGAQQG FLLRDPLDDP DTRQALLELL
 151 GACQEAPRPH LGEFEADPRG QLWQRLWEVQ DGRRLQVGCA QVVPVPEPPL
 201 HPVVPDLPSS VVFPDREAAR AVLEECTSFI PEARAVLDLV DQCPKQIQKG
 251 KFQVVAIEGL DATGKTTVTQ SVADSLKAVL LKSPPSCIGQ WRKIFDDEPT
 301 IIRRAFYSLG NYIVASEIAK ESAKSPVIVD RYWHSTATYA IATEVSGGLQ
 351 HLPPAHHPVY QWPEDLLKPD LILLLTVSPE ERLQRLQGRG MEKTREEAEL
 401 EANSVFRQKV EMSYQRMENP GCHVVDASPS REKVLQTVLS LIQNSFSEP
      (SEQ ID NO: 2)
```

FEATURES:
Functional domains and key regions:

```
PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site
Number of matches: 3
       1      263-265        TGK
       2      275-277        SLK
       3      322-324        SAK
----------------------------------------------------
PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site
Number of matches: 4
       1      271-274        SVAD
       2      378-381        SPEE
       3      394-397        TREE
       4      445-448        SFSE
----------------------------------------------------
PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site
              407-414        RQKVEMSY
----------------------------------------------------
PDOC00008 PS00008 MYRISTYL
N-myristoylation site
Number of matches: 4
       1      21-26   GVCAGA
       2      121-126        GGQAGG
       3      259-264        GLDATG
       4      390-395        GMEKTR
----------------------------------------------------
PDOC00009 PS00009 AMIDATION
Amidation site
Number of matches: 2
       1      17-20   LGRR
       2      181-184        DGRR
----------------------------------------------------
PDOC00017 PS00017 ATP_GTP_A
ATP/GTP-binding site motif A (P-loop)
              259-266        GLDATGKT
```

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	274	294	1.133	Certain
2	489	509	1.775	Certain

BLAST Alignment to Top Hit:
Alignment to top blast hit:

FIG. 2A

```
>CRA|18000004977190 /altid=gi|10181118 /def=ref|NP_065582.1|
            thymidylate kinase family LPS-inducible member;
            thymidylate kinase homologue [Mus musculus] /org=Mus
            musculus /taxon=10090 /dataset=nraa /length=431
          Length = 431

 Score =  262 bits (663), Expect = 1e-68
 Identities = 133/192 (69%), Positives = 152/192 (78%), Gaps = 5/192 (2%)
 Frame = +2

Query: 704  ECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP 883
            +CTSFIPEARAVLDLVDQCPK++QKGKFQV+AIEGLDATGKTT+TQ      +
Sbjct: 157  QCTSFIPEARAVLDLVDQCPKEVQKGKFQVIAIEGLDATGKTTLTQHFKSLSRLSSYSRH 216

Query: 884  PSCIGQWRKIFDDEPTIIR-----RAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATY 1048
            P CI    K  +++  ++        F    NY+VASEIAKES   PVIVDRYWHSTATY
Sbjct: 217  PPCI----KPVEEDLLMMNLLSFEEPFILWANYLVASEIAKESTNFPVIVDRYWHSTATY 272

Query: 1049 AIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAE 1228
            AIATEVSGGLQ+LPPAHHPVYQWP DLLKPDL+LLLTV+ EER++RLQGRG EKT+EEAE
Sbjct: 273  AIATEVSGGLQYLPPAHHPVYQWPGDLLKPDLVLLLTVNSEERVRRLQGRGQEKTKEEAE 332

Query: 1229 LEANSVFRQKVE 1264
            LEAN+VFRQKVE
Sbjct: 333  LEANNVFRQKVE 344 (SEQ ID NO: 4)


SignalP results:

< Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C    49        0.362   0.37   NO
  max. Y    49        0.220   0.34   NO
  max. S     4        0.884   0.88   YES
  mean S     1-48     0.369   0.48   NO
# Most likely cleavage site between pos. 48 and 49: ALG-AD

Hmmer search results (Pfam):
Scores for sequence family classification (score includes all domains):
Model    Description                                    Score    E-value  N
-------- -----------                                    -----    ------- ---
PF02223  Thymidylate kinase                              50    1.6e-12
PF01712  Deoxynucleoside kinase                         4.0          8   1
PF00457  Glycosyl hydrolases family 11                  3.3        8.8   1

Parsed for domains:
Model    Domain  seq-f seq-t    hmm-f hmm-t      score  E-value
-------- ------- ----- -----    ----- -----      -----  -------
PF02223    1/1     257   437 ..     1   197 ..     50  1.6e-12
PF00457    1/1     253   270 ..   177   194 .]    3.3      8.8
PF01712    1/1     359   390 ..    71   100 ..    4.0        8
```

FIG. 2B

```
   1 AAAAGTTAGG AAGAAGCTGG TCTTCCTACT TACCCTCAAG GTGCTCAGTG
  51 GTGGGAAATA GACTGGCACG TGACTGTGGA GTGTCAAGGT CCGAGGATAC
 101 AAATAGCCTT GGTGGGGGAC ACAGGAGAGG CTACTAACCC AGATATGTGG
 151 GGGGTTATAG GCTATATCAT GTCCCCCAAA ATTCATGTTC AGGCCTAGCC
 201 CCCAGTGCCT CACAGTATGA CTGTATTTCA AAATAAGCCT TTCAATAGGT
 251 GATTACATTA AAATGAGAAC ATTATGATGG GCCCTAACCC AATCTGACTG
 301 GTGTCCTTCC CAGAAGAGGA AATTTGGAAC TTCAAGGAGA CGCCAAGGAT
 351 GCAAGCACAA AGGAAAGGTC CCGTGAGAAC ACAGAGAGAA GGAGGCGGTC
 401 TGCACGCTAG AAAGAGAGGC CTCAGAAAAA GCAAACCCTG CTGGCAACTT
 451 GATCTCCAGC TTCTCTGGCC ACCTGAACTT TGAGAAAACA AATTTTCTAT
 501 TGTTTAGGCG GCCCAGTCCG TGGTATTTTG TTATGGCAGC CAGGATAGAC
 551 TAATACATCC AGAGAGGTGG ATGGCATAGG GAAAAGGTCA AGCAGGCTTG
 601 AGGATGCAGT ACTGTCTTAG AGAAGAAAAG GGAGTAACTA GAAGACTCTT
 651 ACTTCCATAT ACTACATACA CGTGAAACCA CCGGTACATG CTAAATGTCC
 701 AAAAGTGAAA TTCCTGAGAA AGAATAAAAA CTATTCTATT CTCAAAGAGC
 751 TAAAAAATTT TAAACTCCTA ATTCTTCTTT CTTCTTAGCC TATTAACCCA
 801 CACTCCAACC ACCTATCTAT ATTTCACGTT TGTTAAGACT TTTTTTTTCA
 851 GAATTAAATA TAATCCCAAG TTTAAGTCCA ATACTATTAA GACATAAAAA
 901 AAAAAACTGC TTTCAGCTCA TCTAATGTTT TCAATCTTCG TCTCAATTCT
 951 ATTTTTTTGA GAGAGTTTCC CTGGAGAATA TTATCTTTTT GTTTGTTTTT
1001 GGTTTGCTTG ACATCATTTT TAAAAGGCAT CAGTTAATGA GTAAACACAG
1051 AATAAAATAT CCAAATAACT GCGCAAACAC TGTTACACTG TTAGGCAGTT
1101 ACACTGTTAG GCAGCAACAG TGCTGATGCT GGACTGTGGC AGGCAGAGGG
1151 TGCTATCCTG ACACACTTCA CCTTAGTGCA GGAAACTTCA ATTTGGTGGA
1201 AGAAAGGCGA TTTCGAGGTT CCAATCTGGG CGACACTTCC CAGTTGGAGA
1251 GTCAGCAAAA GGGAGAGGGC AATTCCAAGA AGAGGGAAAA GCTTGTGCAC
1301 AGGTGAGTGT GTGCAAAGGT GAGTGTGTGC ACAGGTGAGT GAGTGCACAG
1351 GTGAGATAAG ATACAGGAGA GGGAAGAGCC AGAACTGCGC CCTGTTCCCG
1401 CAGGAAAGCA GCTCTGCAGG ATTAGAGGGG CGGGGACGCG TGGTCAGAGC
1451 TAGGAGTTGA GGTCGGGGAG GGAGCCCATG GTCTGCAGGG CCTGGTCAGT
1501 CATCCAAGGG CAGTAGTGCG CCTGCAAGTG GGCGTTGAAG AGCCCGTTAC
1551 ACCGGGAAGG GACTTCTCTG TCCCTCGCGC GTGCACCCCC GCCCCCCTCC
1601 ATGCACCCGG CATAAGCCGC AGAGGAGGAA CTCAAACCAG GGTCGGGGCC
1651 GCCAGCCACC CGCAGAACGC ACACGGAGCT ACCTTGGGGC CGACGGCGCG
1701 GGGCCTCATT CGGTGTCAGC CCCGGGAGCC GGCGCCTGGG GACCGCGCAG
1751 GCCCGCGGAG CCGCGCACCT GGGGCCCCGG GGCCAAGCGT CTGCTCCCGA
1801 GCGCCGGCCG TTTATCGCGC ACATCTCGCG GCGAGGAGGA GAGGCCGGAA
1851 GGGCGCCCCA GCCCCAAGGC TCCTGCCCCG CCTGGGCCTC CGGCTTTCGT
1901 TTCCCCGCAA CGCTTCGCTT TCGTTTCCCG CTGGCGCCTG GCTCCCTCCG
1951 GGTTTCGTTT CCCGCCGGCG CCTGGCTCCC GCCAGGTTTC GTTTCCGAGG
2001 CGGGGCCGAG GGCGGCGTCG CTGAGGCGCC CATGGCCTTC GCCCGCCGGC
2051 TCCTGCGCGG GCCACTGTCG GGGCCGCTGC TCGGGCGGCG CGGGGTCTGC
2101 GCTGGGGCCA TGGCTCCGCC GCGCCGCTTC GTCCTGGAGC TTCCCGACTG
2151 CACCCTGGCT CACTTCGCCC TAGGCGCCGA CGCCCCCGGC GACGCAGACG
2201 CCCCCGACCC CCGCCTGGCG GCGCTGCTGG GGCCCCCGGA GCGCAGCTAC
2251 TCGCTGTGCG TGCCCGTGAC CCCGGACGCC GGCTGCGGGG CCCGGGTCCG
2301 GGCGGCGCGG CTGCACCAGC GCCTGCTGCA CCAGCTGCGC CGCGGCCCCT
2351 TCCAGCGGTG CCAGCTGCTC AGGCTGCTCT GCTACTGCCC GGGCGGCCAG
2401 GCCGGCGGCG CACAGCAAGG CTTCCTGCTG CGCGACCCCC TGGATGACCC
2451 TGACACCCGG CAAGCGCTGC TCGAGCTGCT GGGCGCCTGC CAGGAGGCAC
2501 CACGCCCGCA CTTGGGCGAG TTCGAGGCCG ACCCGCGCGG CCAGCTGTGG
2551 CAGCGCCTCT GGGAGGTGCA AGACGGCAGG CGGCTGCAGG TGGGCTGCGC
2601 ACAGGTCGTG CCCGTCCCGG AGCCCCCGCT GCACCCGGTG GTGCCAGACT
2651 TGCCCAGTTC CGTGGTCTTC CCGGACCGGG AAGCCGCCCG GGCCGTTTTG
2701 GAGGAGGTAA GAGTTCTGTC CGCTTCCAGC TCCCAGCGTG GCATCTGAAC
2751 CCTTCAGACC AGAGAACTGG ACCAAGAGGC TGGTCTGTAA AGCCGGTTCT
2801 TGCCTGGGTG GTTTGTTTAT TTCCGTTCAC AAATCAGGTA GGGAAGGTGT
2851 CCTGTATGCC AGGCAACTCT TTTAAGATTC TTGTTTGCAA GGATCTTCCA
2901 ACCTGACGTG GAACATAGAC CTACACCAAG CCACGCGATG CTTGCTGTAA
2951 AAGCATCCCA ACAGCAGTAC AGAGGGAGTA AAGGGGCTGC CGGGAGTGAG
```

FIG.3A

```
3001 GGAAAATAAT GTCAGCTGGG AAGTAATTTT ATTTGCTGAT GATCACCATT
3051 CAAGGATCTT GGGGTGAAAA AGAAAATGCA TGAGTTTAGG GGGTTTAAGA
3101 AATTTAGACT TAAATAGTGT TTACCTACCG ACTGGCCATG AACCTTGTGC
3151 AGGTTACTCA ACTACTCTAA GTTTTGCCCT TTGACATGTA CAATTCCCAT
3201 CTTGCGATGT TGTCCTGATT AAGGAAACAT CTGACTCACA GCAGGTACCC
3251 ACAGAAAGAG ACTGAAAATT CTTTCTGATC GCAGGCTAGG CAGATGTCCT
3301 CCTGTGACAC AGATGAGCCC TGAGGATGCC CCCATGGATC TTGGGAATAT
3351 TTTCCAAGCT TACGGGACAG CGTTGTGGAG CAGTTAAGAG TGCAAGGTCA
3401 ACCACGTGTA TTTAAATTTA AACTCTGGCA TTTATTAGCT GTGTCACTTT
3451 GAGCAAGTTT CTTCAACTCT CTGCCTCAGT TTCCTTAATT CATATGGTGG
3501 GAATAATAAT AGCACCCCCC CCACCCCAGT TCACAGAGAT TGGCAACTGA
3551 ATACTTGTAA AGCACTTAGA AGATTGCCAT GCTCAGAGCA AGCACATAAG
3601 TGTCTGAGCC TCGCTCTGAG ATGCTGTGAG CGTGCAGTGA GATAATGCAC
3651 ATTGAGGAAC TGGGAATTCC CAGGGGGACG CTGCTCTGCC AGCTTCATGA
3701 TTGCAGTGCT TGGCTGTTTA TCTCAGCCCC CTGAATGGCT AGGAGAGGAC
3751 ATGCTGCAGA TGAAGACTGC TCTCTCCAGC CCACTGTGCA GCTGATTTCC
3801 CATTCTTGTG ACACAGTGTT CCCAGCGGGC CTGTAGTTCC ATGGTTGCGG
3851 TGTCACAGGA CATTGTGATG ATGTGCCTTG CCTGGCCTTT CTCAAAGCTG
3901 CTCAGTGAAG GCTGCAGGCC ACCAAGCGAT CCAGACAGGG ACAGCTGTTT
3951 CGAGCCTTGG CTGCTCAATA TAAAATAAAA TACTCCAGTC CATCCTAGCA
4001 TCGAAATACT CTGAATTCCC ATGGCCTGGC ACAGTGCTCA GTATAACTTA
4051 GGCCTTATTA GCATGCGGCA ATATTGTGCT CAGCAATTTA GGTGTGATTT
4101 CTGCAAAAGC CCCCTGGCTT CATTGCTGAT GGATAGACGT TGTTTTACAG
4151 TGTACCTCCT TTATTCCTGA AGCCCGGGCA GTGCTTGACC TGGTCGACCA
4201 GTGCCCAAAA CAGATCCAGA AAGGAAAGTT CCAGGTTGTT GCCATCGAAG
4251 GACTGGATGC CACGGGTAAG ATAATATTAC CTTTTAGTTA TAGGCAATGA
4301 CACTAACTGA TTAGTTGCAG AAACAGAAAT ACTTCCTGCA AAACCAAACT
4351 TTATATGGAG CCTTATGTGT GCCCCTACTG TGTGGCAGGC CCTGTGCTAG
4401 GCAGGCCCTG GGATGCAGAG ATGAATAAGA CCTTCAATAT GAAGCAGCAT
4451 GATGTGTGGG CGCGGATCCT CAGTGCTCTG GCGGAACACA GGAAGGGCAC
4501 TGAATCTGGC CTCTGTGGGG CTTTGTCGGG TGGAGTGCAT GGTCAAGGGT
4551 GATACCTGGA TTGTATTTTA AGTACAGATA GGAGTTGGTC AGGTGATGAA
4601 AGCAGGTAAC ATCCTCCAGA CAGAAGAAAT AGCCTGGGCA AAGGTGCAGG
4651 GGCTTGAACC AGGGTGGTGT GTCCAGGAAC CACAGGCAAT TCAGAGATTC
4701 TTCTGGAGCA AAATGTGGAA GAACTAGGAA ATGGAAGAAA AAAAAGCCTT
4751 CTGAGCTGTC AAACTGAGGT CAAAATATAA TGTGTGCTCA CATGAGACCA
4801 AAGTACAAAA GGGGCAGACA TGCTGCTCCT GTGGCCCAGG ACACACTGAG
4851 GAGAGGGTTG ATGTTGGAGA ACTAGCATCC GAGTGGTTCA GCGTAGGAGT
4901 TTCTCCTCCT GTGTAAACTT GAGGGGTACA GACTTTTAAT AATATAAAAG
4951 GCAATTTCCA TATAGAGGTA CTTGTGAACC CAGCTAGGGA GATGTGGCAC
5001 AGGTGATGGC CCATGTTGAC CATCCTGGCT CCATGTGAAG GAGCGGGCCA
5051 TGTCCTGGCC TTCAGGGAGA CCAGCTGTCA TCACTCAAAT GTACTGGCCG
5101 TGTCCAGGAC CCATCACAGT TTCTTTCAGC TGCAGAGGGA ATTGTAACAC
5151 CATCAATCCT TCAGCTGATG TGTTTTGTTG ATCATTTATT TTGTACCCAC
5201 AGCTAATTTA GATTTGGTGG GATTACAGGA GACATAAAAA TTCAGCCTCA
5251 ACACAAGCAT CCACACATAC AAATGTTACA AGGAGTTAGC ATAGAGTGGC
5301 AGAAAGAACA CAGGCTAGTG GTTCCGTCTG CCCTGCATTT GCATCCTGCA
5351 CCAAGGCCTA TCAGGAGGGT GAGTTTGGGT AAATTTATTA ACTTCTCCAG
5401 TGCTCAATTT CTTCTTCTGT CAGATGATTA AATAATAACT GTTTTGCAGA
5451 GTCACTGGGA GAATTAGGAG TATAACGTGC TGAGTACTTG GCTCCTAGCA
5501 GACACTGAGA AATGGTAGCT ACTGTTAGGG TCCGTCCTGA CAACCTAAGA
5551 AAAAAAGAAA ATAGATAGTT GGCAATAAAG TGTTAAGTGT GTGATAGAGA
5601 AACTTAAAAA TAAATCGAAA CAGTAGGAGC TCAGAGAAAC TAGTGCACAG
5651 TGTGCTGGAG TAGATCTTCT CATCACCACC TGTCCTGAGC TCCAGGCAGC
5701 AGCTGAGAAT TGTGAGATGG GCTCTGGGAA GGGACTAATC TGTCACCCGA
5751 GGCTGTGCAA GGGGGAGTCA GAAAGTCAAT GAGGCCTAAG CAGTGCCTTT
5801 GAGGAGAAAG CTGAAGCCTA AAGCAGATAC AAAAGCTCTA AAGGCCAAGG
5851 CCAAGCCCAA AGGGGCAAGC ACAGGAGTGA GTGGTAGAAC CAGGGCTGGA
5901 AATTGGAAAG GGATTGCACA GAAGTGGAAG CAGGGTATGA AGAAGGTAGA
5951 AAGAGAGGAG GGGCGAGAAG AGTTGCTGTG GATGCCAGGT GTGGGTTCAT
```

FIG.3B

```
6001 CAACTATAGA CAATAAGAGG AGAGAAAGTC TTCTGGGTTG GGGACATGGT
6051 AAGAGGGTGA GCAGTAGCTG GGCTGCCGGG AATAAAAGTC ACACGTAAAA
6101 GGGGCTCTTG TGTCTAGACT CCCAATATCA GATTTGATCA CTAACCAGAA
6151 TTTTTCCTCC GGGTTTCCTA ATACCACACG GAGAAATCCT AACTTCCTAT
6201 GGGTCTACAG CTTTTTATAA AGAATCCTGT TATTTAGTCT ACTCATTTCA
6251 TTTGCAGTTT GAGAGAACGT CTGTGCTCTT TCTACGTCAA TGTTAACTTT
6301 GGGGCTGTGG TTAAGATGTA TATATTTTGT GTATGACCTG CAGGTAAAAC
6351 CACGGTGACC CAGTCAGTGG CAGATTCACT TAAGGCTGTC CTCTTAAAGT
6401 CACCACCCTC TTGCATTGGC CAGTGGAGGA AGATCTTTGA TGATGAACCA
6451 ACTATCATTA GAAGAGCTTT TTACTCTTTG GGCAATTATA TTGTGGCCTC
6501 CGAAATAGCT AAAGAATCTG CCAAATCTCC TGTGATTGTA GACAGGTAGG
6551 TATAAAGATG CCTTGAATTA GGCATTTTCT CCCTAATATA TAAGTGTGTG
6601 TGTGTGTGTG TGTGTGTGTG TGTATACGTA CATGTATATG CCAGGAAAAA
6651 AATTGTGTTT AAGTCAAACT GTTATTATGA TAATAATAGG AATTCTCCTT
6701 ATGAATTGTT AATTACCTAT ACCAGGCATG GCATTTGCTA GAGAATTACA
6751 TATATAATAC TAGTATCTGG AACTATAACT TGGGTAGGTG AATGTTACAT
6801 GTTATTCCCA GTTTACTGAT GAGAACTATA GATCTCAGAA AGGTAAAATA
6851 ACTTGCCAAG GTAAGCTGGA AATAGCATAC CGGGGACATT AATGAGTCTA
6901 TACTTTCAGT CATTTATTTG TTCATTGGCT TATTCAACAA ATACTTACTG
6951 GACACCTCCT GTGTGCCAAA GACTAGTCTC AATTTAGAGG ATTCAATGAT
7001 AAACCAGTGT ATTAGTCCAT TTTCATGCTG CTGATAAAGA CATACCCGAG
7051 ATTCGGCAAT TTACAAAAGA GAGCAGTTTA ATGGACTTAC AGTTCCATGG
7101 GGCTGGGGAG GCCTCACAAT CATGGTGGAA GGTACAAAGC ATGTCTCACA
7151 TGGCGGCAGA CAAGAGTAGA GAGCATGTGC AGGGAAACTC CCCTTTTTAA
7201 AATCATCAGA CCTTGTGAGA CTTATTCACA ATCATGAGAA CAGCATGGTA
7251 AAGACCTGTC CCCATGATGC AATTACCTCC CACTGGGTCC TTCCCACAAC
7301 ACATGGGAAT TCAAGATGAG ATTTGGGTGG GGACACAGCC AAACCATATA
7351 ACTCCACTCC TGGCCTCTCC CAAATCTCAT GTCCTCACAT TTCAAAACCA
7401 ATCATGCCTT ACCAACAGTC CCACAACTCT TAACTCATTT CAGCATTAAC
7451 TCAAAAGTCC ACAGCCAAAG TCTCATCTGA GACAAGGTAA GTCTCTTCCA
7501 CCTATGAGCC TGTAAAATCA AAAGCAAGTT AGTTATTTCC TAGATACAGT
7551 GAGGCTACAG GCATTGGGTA AATACAGCCA TTACAAATGA GAGAAATTGA
7601 CCAAAACAAA GGGGCTACAG GCTCCATGCA AGTCTGAAAT TCAGCTGGGC
7651 AGTCAAATCT TAAAGCTCCA AAATTATCTC CTTTGACTCT ATTTCTCATG
7701 TCCAGATCAT GCTGATGCAA GAGGTGGGTC CTCATGGTCT TGGACAGCTC
7751 CATCCCTGTG GCTTTGCAGG GTATAGCCCC CCTCCTTGCT GCTTTCACAG
7801 GCTGGTGTTG TCTGCAGCTT TTCCAGGTGC ATGGTGCAAG CTGTCAGTGG
7851 ATCTACCATT CTGGGGTCTA GTGGACAGTG GCTCTCTTCA AACAGCTCCG
7901 CTAGGTAGTG CCCCAGTAGG GACTCTGTGT TGGGGCTCCA ACCCCACATT
7951 TCCCTTCCAC ACTGCCCTAG CAGAGGTTCT CCATGAGAGC CCCACTCCTG
8001 TAGCAAACTT CTGCCTGGAC ATCCAGGCAT TTCCATACAT CCTCTGAAAT
8051 CTAGGCGGAG GTTCCCAAAC CTCCATTCTT GACTTCTGTG TACCTGTAGG
8101 CTCAACACCA CATGGAAGCT GCCGAGGCTT GGGGCTTTCC CCCTCTGAAT
8151 CAAGAGCCTG AGCTGTACCT TGGCCTCTTA CTCAAGGCTA GAGTGGCTGG
8201 GACACAGGGC ACCAAGTCTC TAGGCTGCAC AGAGCAGAGG GACCCTGGGT
8251 CCACAAAACC ATTTTTTTCC TTCTAAACCT CTGGGTCTGT GATGGGAGGG
8301 GCAGCAGCAG AGGTCTCTGA CATGCCCTCG AGACATTTTC CCCATTGTCT
8351 TGGTGATTAA CATTTGGCTT CTCATTGCTT ATGCAAACTT CTGCAGCCAG
8401 CTTGAATTTC TCCTCAGAAA ATGGGATTTT CTTTTCTGTC ACATTGTCAG
8451 GCTGCAAATC TTCCAAACTT TTATGCTCTG TTTCCATTTT AAAACCGAAT
8501 ACCTTTAACA GCATCCAAGT CACCTCTTGA ATGCTTTGCT GCTTAGAAAT
8551 TTCTTCCACC AGTTACCCTA AATTATTCTC TCAAGTTCAA AGTTCCACAA
8601 ATCTCTAGGG CAAGGGCTAA ATGCCGCCAG TCTCTTTGCT AAAGCATAAC
8651 AAGAGTTACC TTTGCTCCAG TTCTCACCAA GTTCCTCATT TCCATCTGAG
8701 ACCACCTCAG CCTGGATTTC ATTGTCCATA TCATTATCAG CATTTTGGTC
8751 AAAGCCATTG AACAAATCTC TAGGGAGTTC AACCTTTCCC ACATTTTCCT
8801 GTCTTCTTCT AAGCCCTCCA GACTGCTTCA ACCTCTGTCT ATTACCCAGT
8851 TCCAAAGTTG CTTCCACATT TTTGGGTATC TTTTCAGCAG CACCCCACTT
8901 CTGGTACCAA TTTACTGTAC TGGTTCATTT TCACACTGCT GATAAAGACG
8951 TACACGAGAC TGGGCAATTT ACAAAAGAAA GAGGTTTAAT GGATTTACAG
```

FIG.3C

```
 9001 TTCCATGTCG CTGAGGAGGT CTCACAGTCA TGGTGGAAGT TACGGCACAT
 9051 CTCACATGGC AGCAGACAAG AGTAGAGAGC TTGTGTAGGG AAACTCCCCT
 9101 TTTTAAAACC ATCAGATCTT GTGAGACTTA GTCACTATCA TGAGAACAGC
 9151 ATGGGAAAGA CCTGCCCCTG TGATTCAATT ACCTCCCACT GGGTCCCTCC
 9201 AACAACATGT GGGAATTCAA GATGAGATTT GAGTGGGATC ACAGCCAAAC
 9251 CATATCAATG AGATAGATAA GTCCCTATTT TCATGGAGCA AACTTAACAT
 9301 TATAGGAGAA GAAAAGTATC AGGTGAACAA ATACATAAAA TAATACATAA
 9351 GATGAGGTAA GATAATATCA AAGCATGATA AATGCAGGGA AGAGGAAAAA
 9401 TCAAAGTAAT GTGCTAAAAA ACGGCTAACC CTCCACTAGA TATGGTTTAG
 9451 GAAGGCCTGT CTGAGAAAGC ACCATTAGTC AGAGCCCTGA TTTAAAAAAA
 9501 AAAAAGGCAA ATGTGAAAAT TCCCGGGTTA ACAGAAAGCA CTGTGGAGAA
 9551 AGAAATCTGC AAGAATGAAG CTAAGACTGA AATAAGCTAA CATATCTGAC
 9601 AACTAGAAAA TGTTATATGT TCTGAGAACA TAGTAGATGT GGAGGTGCTT
 9651 TGTGGATGAA TGGGAAGAGG AAGGTTGGGG CAGGTCTGTA GGGCTTGTAG
 9701 GCCATTCATA GAATGGATTT TATTCTGAGT GCACTGGGGA GCCATTGGAA
 9751 TGTTTCTGAT AAAGGAGAGA CATAAACTGA TTTATACTTT AAAAATTCAC
 9801 CTGTAAGAAA TAGCTTCACT TTGGGAGGCT GAGGTGGGCG GATCATGAGG
 9851 TCAGGAGATC GAGACCATCC TGGCTAACAC GGTGAAACCC CGTCTCTACT
 9901 AAAAATACAA AAAATTAGCC GGGCGAGGTG GTGGGCACTT GTAGTCCCAG
 9951 CTACTCGGGA GGCTGAGGCA GGAGAATGGC ATGAACCCCA GGGGGTGGAG
10001 CCTGCAGTGA GCCGAGATCA TGCCACTGCA CTCCAGCCTG GGCGACAGCG
10051 AGACTCCGTC TCAAAAAAAA AAAAAAAAGA AAAGAAAAGA AATAGCTTTA
10101 GGTCAGGTGC AGTGGCTCAC ACCTGTAATC CAAGCACTTT GGAAGGCTGA
10151 GGTGGGAAGA GTGCTTGAGC CCAAGAGTTC AAGACCAGCC TTGATAACAT
10201 AGTGAGACCT TGTCTCTACA AAAAATATTT AAAAAAATAG CTGGGTATGG
10251 TGGCACCCAC TTGTAGTCCC AGCTACTTAG GAGACTCAGG AAGGAGAATC
10301 CCTTAAGCCC AGGAGGTCGA GGCTGCAATG ACAAAGCAAG AGGATGACAC
10351 AGCAAGAGGG GCCTTGTCTC AAAAAAAAAA AAAAATCCC AAAAACAATA
10401 AAACAAAAAA CAAAAAAGAA GAAAAAAATA GCTTTAAAAA TACTATAAGA
10451 AAGAGGAAAG GATATGACAA GCAGGTCATA CTGAACTTAC ATCACGATGA
10501 CAAATGGATG AGAAGATTAT AGTCACCATG AGATTCCATT TTACTCCCAT
10551 CTGGATTTCA TTTTACTCCC CTCCAGTTGG TGAAACTACT ATAACTAAAG
10601 GTTTCTTCAT ATGATATGGA TCAATAGAAA CTCTCATATG ATGCTAGTCA
10651 GAACATATCA TATGATAGTA GCATAATCAC TTAGAGAGCA GTTTGAAAAG
10701 ATGTAGCAAA GTGGAAGATT GCTCCTGCCT ACTGAAGAAA CACTCACAAA
10751 AGTTTCTTTC AGCTTCCTTC ATACTGTTGC AAAATTTGAA AAACCTAAAG
10801 CCCATCAACA GGAGAAAGCA TAAATAAATA CCAGCATATG TATAAGGTGG
10851 AAAACCATAC AGCAGCAGTG TAAATATGTT GTGCAATGCA TGAGGGCACA
10901 CACATTACAA AACCGTAATG TTGAATGAAA ATGTCAAGTT GCCAAAAGAT
10951 GTTTATTCAG TAGAATACAT GCAGTTATCC ATACTGCAAA GAGTAGAGAG
11001 ATGAAAGGAT GTTGAACAGC AGATTTAAGA AAGTTGTTCT CTGGGAAGTA
11051 GAGATAAATG TGGTTGATAA GGAGATCAGA AGGACTTTGA TTGTGATTGG
11101 ATTATTTCTT AACATTTCTT GAATGTTCAT TTTCTCCAAA GATTTACTTA
11151 TTTTTGTAAA CATAAGCAAT TGCATTTAAG TATGGCATAG TTGGTAAAGA
11201 ACATTAATCA TAAATGTAAA TCTCAAGGAA TTTCAAGAAG TGAACACAGT
11251 CAAGTCACCT GTACCCAGAT CAAGAATGAG AGCCTGCCAC AGCCCCTACA
11301 CCCCTGTTCT GCCCCCACCC AGTCACTGCC CATTTCCTCC CCAGGGGTAG
11351 CAACTGTCCC AACTGCCATA GCCATATATT AGTTCTGCCT GTTCTTAAAC
11401 TGCCTACAGA TGCAACTATT TGTTAGGTAT GGAGTTTTAT TTGCACCTGA
11451 CTTCTTTCAT ACCACATGTG TTTTGAAGAT TTGCCTGTGT TGTTGTGAAC
11501 TAAATTTTAT GCATTTTTGG TGCACATATG GATGCAGTTC TGTTACATAG
11551 GTTCCCAGGA GCAGAACTGC AGGGGCATGG GGTTTGTGGC TGTTCAGGTT
11601 TGGTAGATGT TGCAAAAAGT ATTCTAAAGA GGTTGAAAGA ATGTTGAGTC
11651 TCCTCCCCTC CCCACAGCAG TATTTGCCAA TTCCCTGTGA CTAACAGCCC
11701 TGTTGACATT AGTATTACCA GGCTTTAATT TTGACTCCTC TAGTTTTTCC
11751 ACTTGTTCCT GGTGGGTAAG CTGGTCACTG AAAAGCTGAT TTGTCCTATC
11801 TGGACACAGA ACTACTACTC CTTTCTTGAA AGTATATCTT TTTCTGTACA
11851 CTTTTCTGTA TGTCTGAAAT ATTCCATCAA GAGACCACCC TGACTGTATG
11901 TGGAGAATAA AGTGTGTGGT GGGGGAGGGG CAAGAAAGGA AAAAGAATCA
11951 GCTAGAAGGC TGTGGCTGCA GTCTAGGCTG TGATGGCCCT TGAGTAGTTT
```

FIG.3D

```
12001 TAATCCTGGG ATGACAATAG GGATGGCAAT GGGGATAACA ATAATAATTT
12051 CAAGGTTGGC AAGTTACATG GAATCTTAGG ATGAACTGAG AAGGATACGA
12101 AATCTTGTTC CTTTATGGGA AGGCTTTGCC TGGAAAATGT TTTTGCCCTA
12151 TTGTTAACAT GCCCGATGTT ATCATATGTG ATATCTGTAT TAGTTCTTTC
12201 TCACACTGCT ATAAAGACGT ACCTGATACT GGGTAATCTA TAAAGAAAAG
12251 AGGTTTAGTT GACTGACAAT TCTGCATGGC TATACGGGAG GCATGGCTGG
12301 GGAGGCCTCA GGAAACTTGC AATCATGGTG GAAATTGAAG AGGAAACAAT
12351 GGCAACAGGA GAGAGAGAGA GAGAGGCAGG GGCAGGGGCA GGGGCAGAGG
12401 CGAGAGAGCA GGGGGAAGTG CTACACATTT TAAACACCCA GATCTCATGA
12451 CAACTCACCA TCACGAGAAC AGCAAGGGGG GTGTCTGCCC CCATGATCCA
12501 ATCAACTCCC ACAGGGCCTC ATCTCCAACA ATGGGAATTA TTAGGTTGGT
12551 GCAAAAGTCA TTGTGGTTTT TGATATTATT ATTAATGGCA AAAACCGCAA
12601 TGACTTTTGC ACCAACCTAA TACCCTTTGG CATGAGATTT TGGTGGGGAC
12651 AAAGAACCAA ACCATATTGA CATCTTTTTT GATACAGTCC CCTTTATTTC
12701 CAAGAGAAAG ACTAAGGTTT TCCTAGTAGG GTATGACTTT CGAGGTCCAT
12751 TATGTCCTAG GATGCCTGCG GATCCTGAAG CAGCACTGGC CACTGTGTGC
12801 AGGCAGGGAT TTCTGCACTC TGTCCCCCAG TTTTCTGATC TGTTAAGTGG
12851 GGATACTCAT GCCCCCTTCC CTGCCTACTG TCAAGAGTTG TGATGATGAT
12901 TCAGTGAGTG GATGTGTGTG GAAGTACCCT GAAAATAGGA AATTGCTATT
12951 AAAATATAAA GCATTATTAT ATGAGCAGTG TATAATGTGT TGGCAAATTG
13001 CTTTTGATTT GAACTAATGT GGCTTCCCTG ATAGCAGGAG TGGAGAATAC
13051 TAAATAGTGG GAAGCATCTG AAATTGATGG GCTAAGGGTG CAATTATTTA
13101 AAACAAGCAG CCGTATTTTC AATGGGAGAA CTCTATAGGA AACAGGTCCT
13151 TAATTCTTCC CTTGATTTGT CTTCTTTGTG TGTGTGAATT GCCTGCAATT
13201 TAGTTCTTTA AAGAAATGCT GTATCACCTT GTCAGATGAA AAGAAAAGAG
13251 CAGTTATTTG TTGTCTTTGT GGATTTTATT CATGTTTAAA GATTTTAATA
13301 AAATCCATTT TAGACAGTAC CATTATCTAG CTGAAAAATA TGAGAGACAG
13351 TAATTTTTAA CGGGGACTGT GGTTAAGGTT GGAGTCTTAA TCACCCCATT
13401 ACCTTTAAAA ATCTATTCTT GCTGGTGATT TTTCTACAAT AAAGAAGACT
13451 TTAAAAATAA GATAATATCA GACTCTATAT TCATAGGTAG GTATTTAATT
13501 CAATGAATCT GGAGCATGTG CTGACCATGG TGTAAATTAT AGTTTAAGTA
13551 CCAGAAAAAG AAAACTGAGA CCCTAATTGG CTTTTTTTGA GCTTGAGGGA
13601 CAAAATTCAT CTGGCAGAGA GAGTGAAAGT ACAAGTTTGT GAGTAACAGG
13651 AGTTGGGTAA GTAACACATA GGAAGGTGTC CAGGCAGAAT TCACAGGAGC
13701 TGGCAGTGGC CTGAAGCTCT CAGAGCACAC TTTTGGAGGT GAACAAGGGC
13751 TTTGAAGGAT GGATGGTGTT GAGATTATCA ACTCCCAAGT GAATTTTTCT
13801 TTTTTTTTTT TAGATGAAGT CTCGCTCTGT TGCCCAGGCT GGAGTGCAGT
13851 GGCGCGATCT CGGCACACTG CAAGCTCTGC CTTCCGGGTT CACGCCATTC
13901 TCCTGCCTCA GCCTCCCGAG TAGCTGGGAC TACAGGTGCC CGCCACCACG
13951 CCCGGCTAAT TTTTTGTATA TTTAGTAAAG ACGGGGTTTC ACCATGTTAG
14001 CCAGGATGGT CTCGATCTCC TGACCTCGTG ATCCACCCAC CTCAGCCTCC
14051 CAACCAAGTG AATTTTTTAC TTGTGTCCTT TTCAGTGCTG TCCTGTGTTC
14101 TGTTATCATA ATTTGCAATG ATCCGGCTTT AGTTATAACC AGTGTCTGAT
14151 AAGAATTAGA TATTTATCTT ATAGTAACAG TGTGATACAG TTTTTTTTAA
14201 GCACTTGTCT GTATTTGTAA CAACTATGGA AGGAAAACAA ACCTTGCATG
14251 ATCTGTGTTT TCCAGATGAG GAGATGGAGG CTATATTAGC TTAGATGACT
14301 TTTACCTACA TGTACAAAAC AGGTGGGGCG GGGGACACAG GCAGAATAAT
14351 GTACAGTTCA GGTAACACAG GGAATTTATT ATGTGGATAC CACTGTGTAC
14401 TTTTCACTGT GGAGAGGAGT TCAATTCTAA AATGATCAAA ATTTTAGGAT
14451 TTTAAAGAAT TGGGCCGGGC ATGGTACCTC ACGCCTGTAA TTCCAGCACT
14501 TTGAGAGGCC AAGGCGGGTG GATCACCTGA GGTCAGGAGT TTGAGACCAG
14551 CCTGGCCAAC ATGATGAAAC CCCATCTCTA TTAAAAATAC AAAAATTAGC
14601 CGGGCGTGGT GGTGCATGCC TGTAATCCCA GCTACTTGGG AGGGTGGGGC
14651 AGGAGAAATG CTTGAACCTG GGAAGCAAAG GTTGCAGTGA GCTGAGATCG
14701 TGCCACTGCA CTCCAGCCTG GGCAACAGAG TGAGACTCCA CCTCAAAACA
14751 AAAATAAAGG ATTGAAGGTG GTAATTTGAA AGTACAAATG AGGAGGGGCC
14801 CCTGGGTATC TCTACGTTGG AATGTTTATA TCATAAATAT TTATTGTGAG
14851 TGATGGTCCT TTTATATTGG ATCTGAATTG TCCATTTAGT CCTTTAAAAT
14901 TGGAAGATGG CATGAACAGG GCAAGAGTAT AATAAACTAT GCTGATAAAT
14951 GAAATCGTTC TAATTCATTT ATTCATTTAT ACACCCAAAT AACATTCTTT
```

FIG.3E

```
15001 CATTGCATAT CTATTATGTG CCAGACATGC ATTACTAAAA AAGATTTCCC
15051 ACTCAAGAAA TTACCTAATG GAGGAGACAA TAGATATTAC ACTTATAACA
15101 AGTAGTTCTG ATTTGCAATG AGGAGTTTTT CAGTAGAGTC ACGTAGAAAG
15151 GGTTTTGAGA GCACAGAGGA GAAAACAGTC AATTCTGTTT GGGGGCTTCC
15201 TAGGAGTCTG AAGGGAGAGG AGAGTTTTGC TGGCTGAGAA CCTCACTCTC
15251 ACCAGAGGAA AAGGTAAGCA GGTGCAGACT AGGAGGGATG TTCTGTGTCT
15301 TACCAGAAGA CCAAGCATTT CCTGTAGGTT GTAGGAAGCC ACTAGGCATT
15351 TTTAAATAGA GACTGATTTG ACTTTTGTGT ATGGTAATAA CTTTGTTTTC
15401 CTCCCCCAAA ATCACTTTTT AAAACAGCCA GCAGAAGGAG CGACTGACTT
15451 GCTTAGGGAG GACTTTCATG GAGCTGGGCA GGGCATATTT GTCTCCCTGT
15501 CTCATACTGA GGCACCATCA GCAGACTGGA TAGTTGGGAG AAACAAAGAG
15551 GCTTCTACCT CAGGGGTCCC AGAATGTGAA TTTCATTGGC AAGTTCAAGT
15601 GAAAACAGTG TAGGAACTGA CATGGCCTTT CCAGGATTTT AGTCTGCCAA
15651 GACACAGCCC TTAAATACAA ATGAACTGCC AAACAGGTTC ATTGTCCCCT
15701 GTCACCCTCC ATTCTTTCAT AGAGGAATGC GGACAGCAGG ACCAAAAAGA
15751 TGTGATGACA GAGGGGAGGC CACACTAAAT GGTAGTTTGA GATGGGTCAA
15801 TGGAGCTGTG TGAAGAACAC ACTGCATTAT TACTGTTGTC AATTTTATTT
15851 TTTAAACAAT ATTGTATAAC TTTTTTTAGT TTATAAAATT TAATTTTATT
15901 TAACTTATGC AATACTAGAA AAACTTCTGT AGAGCCAACC CTGGTTTCAT
15951 CATTCCTGGC TGCTGATTTT CAGATGCTAC TTTGACTTCT TCTGCAGACA
16001 GAGAACTCAC TACCCCATTC CTTCTCAGGT ACTGGCACAG CACGGCCACC
16051 TATGCCATAG CCACTGAGGT GAGTGGGGGT CTCCAGCACC TGCCCCCAGC
16101 CCATCACCCT GTGTACCAGT GGCCAGAGGA CCTGCTCAAA CCTGACCTTA
16151 TCCTGCTGCT CACTGTGAGT CCTGAGGAGA GGTTGCAGAG GCTGCAGGGC
16201 CGGGGCATGG AGAAGACCAG GGAAGAAGCA GAACTTGAGG CCAACAGTGT
16251 GTTTCGTCAA AAGTAGGTGT CCCAGTGCAA TGCAATGTGA GCGGCAGGCA
16301 TTCCTGAAGG GAGATGAACC ACTGGCACTG GCTTTAGGAT TGTGAGGAAG
16351 TGATATTGTT TCCAGTTTTC AAACACAAGA GACAACATCC TCTAAGTTAC
16401 TTCAGCCCCT TCCAATGGGC TTGTCACCAC AGGGCTGCAG CATTGTTATC
16451 TTAAAGCAAA GGTCATCGGA CTAGGGATCA GACCCTGCCA CTGATCCTGG
16501 CTGTGCTAGG AGCAGCTGCA CCTGGGTAAG ACAGTAAGTG TCTCTGTGCC
16551 TCAGTTTCCC CAGTCATAGT ATAATCACAC AGAGCACTAG ATAACGAGCT
16601 CATAGTAACA TCTACCTATT AGATGCTTCC CGTGTGTCAG GCATTTTACT
16651 GATGTTATGT CATCCTTGTG AGGAAAACAT TAGCCGTATT TTACAGTTTA
16701 CAACTTTAAG GCTCAAAGGA TTAAGTGATT TGTCTAAATG TACATAACTA
16751 TTCACTAGTA AAACCGGGAT TAAAATCTTT CTGATTTTGC AGCCAGTGTT
16801 TTTGTTTTAA TTAGAAAGTT ATAAACACGA CTGCAGAAGA GAGTCTGGCC
16851 AGGCCTCCTG CCTCATGACT GAGTATGAAT CAGTTCTACA CCACTGCCTT
16901 TAAAAACTGA AGCAGAAATA TTTTCTCTAA CTGAACAATG ATAGCCCTGT
16951 TATCATAACA TAGTAATGTT ATAAATAATG GTAGCTGCTG TGGGTAAAGA
17001 TATTATGTTA AGCAATTTAC TTGTATTAAT TCCATTAAAC TTCAGTGAAC
17051 ATTTGCAAGG AAGGTACGGT TTCAGTCTTC ATTTTGCAGA ACAGGAAACT
17101 GAGACACAGA GAGGAGAAGA GATTTGACCA ATTCACTTGG CTAGGAAGTG
17151 GTCAGGCAGA GTTGTGAATG CAGACGATCC ACCTCGACAC CCCTATTTTA
17201 ACCACAGTGC TATAAGGATT CCATAAAGAA ACAGGCACTA GTCACTCTGT
17251 ATACACATGA AGGCTAGCTA GCATGGAAAG GATATGTAGA TTTCTGGCAA
17301 AATATTAGAA GAGTCCCATG CATATATTAA GGACTGTGGC TTGTATGAAA
17351 ATTATTCAGG GCACAGACTT GGGGGAAATT TGCCACTGAA CAAATTACCT
17401 AAATTCTTTG AGCTTCAGTT GCCTTCTCTG CAAAACAGGG ATGACAATAG
17451 TCTTCCCTCC TAAGGTTACT TTGAGAATTA AATGAGAAAA ATCATGCAAA
17501 ATGCTAATGC TTGGCAGAAA ACAGGTATTC AACAAGTGCT AGCTATTAAA
17551 CATTATTATT CATTTATTAT TTGTTAAACA AAATGCATGA ACGTCTTCTG
17601 TGGGCAAAGC TAGTTACAGT GAGATAAATG ACATGGGAAG CTTGCTTCAA
17651 GCTATTTATA CTATGGTAGG AAAAGAACAT TAATGCAAAT AGCTGTGTGA
17701 AAAAGTAGAC CAACTCTTTG TGTTTTGCCT GTCCCCAGGG TAGAAATGTC
17751 CTACCAGCGG ATGGAGAATC CTGGCTGCCA TGTGGTTGAT GCCAGCCCCT
17801 CCAGAGAAAA GGTCCTGCAG ACGGTATTAA GCCTAATCCA GAATAGTTTT
17851 AGTGAACCGT AGTTACTCTG GCCAGGTGCC ACGTCTAACT AGATTAGATG
17901 TTGTTTGAAA CATCTACATC CACCATTTGT TATGCAGTGT TCCCAAATTT
17951 CTGTTCTACA AGCATGTTGT GTGGCAGAAA ACTGGAGACC AGGCATCTTA
```

FIG.3F

```
18001 ATTTTACTTC AGCCATCGTA CCCTCTTTCT GACTGATGGA CCCGTCATCA
18051 CAAAGGTCCC TCTCATCATG TTCCAGTGAG AGGCCAGCGA TTGCTTTCTT
18101 CCTGGCATAG TAAACATTTT CTTGGAACAT ATGTTTCACT TAATCACTAC
18151 CAAATATCTG GAAGACCTGT CTTACTCAGA CAGCACCAGG TGTACAGAAG
18201 CAGCAGACAA GATCTTCCAG ATCAGCAGGG AGACCCCGGA GCCTCTGCTT
18251 CTCCTACACT GGCATGCTGA TGAGATCGTG ACATGCCCAC ATTGGCTTCT
18301 TCCACATCTG GTTGCACTCG TCATGATGGG CTCGCTGCAT CTCCCTCAGT
18351 CCCAAATTCT AGAGCCAAGT GTTCCTGCAG AGGCTGTCTA TGTGTCCTGG
18401 CTGCCCAAGG ACACTCCTGC AGAGCCATTT TTGGGTAAGG AACACTTACA
18451 AAGAAGGCAT TGATCTTGTG TCTGAGGCTC AGAGCCCTTT TGATAGGCTT
18501 CTGAGTCATA TATAAAGACA TTCAAGCCAA GATGCTCCAA CTGCAAATAT
18551 ACCAACCTTC TCTGAATTAT ATTTTGCTTA TTTATATTTC TTTTCTTTTT
18601 TTCTAAAGTA TGGCTCTGAA TAGAATGCAC ATTTTCCATT GAACTGGATG
18651 CATTTCATTT AGCCAATCCA GTAATTTATT TATATTAATC TATACATAAT
18701 ATGTTTCCTC AGCATAGGAG CTATGATTCA TTAATTAAAA GTGGAGTCAA
18751 AACGCTAAAT GCAATGTTTG TTGTGTATTT TCATTACACA AACTTAATTT
18801 GTCTTGTTAA ATAAGTACAG TGGATCTTGG AGTGGGATTT CTTGGTAAAT
18851 TATCTTGCAC TTGAATGTCT CATGATTACA TATGAAATCG CTTTGACATA
18901 TCTTTAGACA GAAAAAAGTA GCTGAGTGAG GGGGAAATTA TAGAGCTGTG
18951 TGACTTTAGG GAGTAGGTTG AACCAGGTGA TTACCTAAAA TTCCTTCCAG
19001 TTCAAAGGCA GATAAATCTG TAAATTATTT TATCCTATCT ACCATTTCTT
19051 AAGAAGACAT TACTCCAAAA TAATTAAATT TAAGGCTTTA TCAGGTCTGC
19101 ATATAGAATC TTAAATTCTA ATAAAGTTTC ATGTTAATGT CATAGGATTT
19151 TTAAAAGAGC TATAGGTAAT TTCTGTATAA TATGTGTATA TTAAAATGTA
19201 ATTGATTTCA GTTGAAAGTA TTTTAAAGCT GATAAATAGC ATTAGGGTTC
19251 TTTGCAATGT GGTATCTAGC TGTATTATTG GTTTTATTTA CTTTAAACAT
19301 TTTGAAAAGC TTATACTGGC AGCCTAGAAA AACAAACAAT TAATGTATCT
19351 TTATGTCCCT GGCACATGAA TAAACTTTGC TGTGGTTTAC TAATCTATGC
19401 TGTCATCCTG GGTACATATT GATTTGTCTG AAAAGTGCTT TCTCAGATTC
19451 CCCTTTTAAT ATTGTGATGT AAAGGAGGGA AATTTTGGTA AAGGAAGTTG
19501 AAAGGTGTGA GCTGGCAGGC TAAGTGGAAT TTGTGGTCAG AGTGCTTTCA
19551 GAGAAAGGGG AGGGCTATTG TTTTATTTTA CATATCATTT CCTCATTACA
19601 AATATTAAAG ACATTTTGTA ATTCATTCTT TTTACACCTG GACTTTTTAT
19651 ATACTGATAG GTATATATGA CTTACGAGTA TTTTGTAAAA TAGCACCTCC
19701 TACCCTAAAA CTGATGGCAA GTAACCCTTT GCTTGGCTCT GCTCATTGCA
19751 AGACGAGCTT TGGTTTTGTT CCTGTGATAG ACCATTAGTT ACCCCAAAAT
19801 TTATTCTTCC TTTCTTCCAT GGTAATGTAA TATTTAGCTT GGTACATGGG
19851 TGCCAATAAT TCCTAGTGCA TTTCCAAGGC TCCCTTAACA GCTGGAGGTG
19901 GGCGACTGGC TAGTTTCTTG CCAATAGTAT GTGAGCAGAA GGAATACCTG
19951 AAATGTCAAG GGGATATTAT CACTTTCCCC TTTAACTCTT CATATCGGCT
20001 GTATTCAGAG GTGATAGCCA TCTAAGGACC AAGAGATGAG GATGACACTA
20051 AAAGGAATTA GCTTGCCTTA GGTAAATAGC AAGGGAAGGG TCCCCAGAGA
20101 GCCCTCGGCC CACCAGTCAG TGCCTCACCC CACATAATGT AAAAAGCAGC
20151 CTGGGGAAAA AATCAAGCTG CAGGCACTGA TAAGGGAACT AGCACAGGGT
20201 GTTGTGCCTG GAGACATGCC TACGGCTGCA CAGATAGGAG AGCCTCTGGC
20251 CCATTCAGAT AAAAGCTTGC ACAAACCTCT GGCTCACTCA GATGAGGGAA
20301 CAAGTCCTGG CATAAAAACA CCTTTGTCCT TTGTATAGTC AGCAGGCTCC
20351 CAGGAAAAAG TTTTTTTCTC CTTTTGTGGG CGTGGGCACA GTGGGCTCCA
20401 GTTAGTTCCA GTGGGCACTT TCCTTGCCAG TTTTTGGACT GTGAGTCTGG
20451 CCTCTGTGAA TCATAACTTC AGCCCCTGAT TGGTCCCAGG CAAAGGTCCT
20501 AGGCCAGGCT TTCTGATTGG TCCTGGGCCA GGGTGCCTGG CCAAGCTGAG
20551 TCATGCCTTC TCCAAGACAG CCGGTAGACT AAGCACATTC ATTCCCCTTT
20601 CCAGCCCGTA AAACCCCCAC AACTGGCCTC ATAGTGGGCA CCCCATTAGA
20651 GCTCCCCCTT CTGCTGGCAG AGAGCTTTCT TTTTTCGCTT ATTAAAGTTT
20701 CACTCCAACT TCACCCTTGT TGTCTGCACT CCTTAATCTT CTTGGAATTA
20751 GGACAAAGAA CTCTGGATAT TATCTCAGAC AACGGGAGAC TGTTACATCT
20801 TGGTGCATTG GTAAGATTAC AACACATTTT GGTGCATTGG CTGGGAAGAA
20851 GGGAATTCAT CAGAAGGATG ATTAAGAGTT GACCTTTAAC TTTCACCTTT
20901 ACTTGCATTT CTGAGGCTTC TTGTCTATTC CAGTCTAGTT TGCTTTCACA
20951 GAGGGCCTAG CCATCA
```

FIG.3G

(SEQ ID NO: 3)

FEATURES:

Start: 2032
Exon: 2032-2706
Exon: 4151-4265
Exon: 6344-6545
Exon: 16029-16262
Exon: 17739-17859
Stop: 17860

Sim4 results:

Exon: 2001-2706, (Transcript Position: 1-706)
Exon: 4151-4265, (Transcript Position: 707-821)
Exon: 6344-6545, (Transcript Position: 822-1023)
Exon: 16029-16262, (Transcript Position: 1024-1257)
Exon: 17739-18966, (Transcript Position: 1258-2491)

SNPs:

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
4530	G	C	Intron			
5856	C	T	Intron			
6594	-	G	Intron			
10987	C	T	Intron			

Context:

DNA Position

4530 TCCAGGTTGTTGCCATCGAAGGACTGGATGCCACGGGTAAGATAATATTACCTTTTAGTT
ATAGGCAATGACACTAACTGATTAGTTGCAGAAACAGAAATACTTCCTGCAAAACCAAAC
TTTATATGGAGCCTTATGTGTGCCCCTACTGTGTGGCAGGCCCTGTGCTAGGCAGGCCCT
GGGATGCAGAGATGAATAAGACCTTCAATATGAAGCAGCATGATGTGTGGGCGCGGATCC
TCAGTGCTCTGGCGGAACACAGGAAGGGCACTGAATCTGGCCTCTGTGGGGCTTTGTCGG
[G,C]
TGGAGTGCATGGTCAAGGGTGATACCTGGATTGTATTTTAAGTACAGATAGGAGTTGGTC
AGGTGATGAAAGCAGGTAACATCCTCCAGACAGAAGAAATAGCCTGGGCAAAGGTGCAGG
GGCTTGAACCAGGGTGGTGTGTCCAGGAACCACAGGCAATTCAGAGATTCTTCTGGAGCA
AAATGTGGAAGAACTAGGAAATGGAAGAAAAAAAAGCCTTCTGAGCTGTCAAACTGAGGT
CAAAATATAATGTGTGCTCACATGAGACCAAAGTACAAAAGGGGCAGACATGCTGCTCCT

5856 AGAAAATAGATAGTTGGCAATAAAGTGTTAAGTGTGTGATAGAGAAACTTAAAAATAAAT
CGAAACAGTAGGAGCTCAGAGAAACTAGTGCACAGTGTGCTGGAGTAGATCTTCTCATCA
CCACCTGTCCTGAGCTCCAGGCAGCAGCTGAGAATTGTGAGATGGGCTCTGGGAAGGGAC
TAATCTGTCACCCGAGGCTGTGCAAGGGGGAGTCAGAAAGTCAATGAGGCCTAAGCAGTG
CCTTTGAGGAGAAAGCTGAAGCCTAAAGCAGATACAAAAGCTCTAAAGGCCAAGGCCAAG
[C,T]
CCAAAGGGGCAAGCACAGGAGTGAGTGGTAGAACCAGGGCTGGAAATTGGAAAGGGATTG
CACAGAAGTGGAAGCAGGGTATGAAGAAGGTAGAAAGAGAGGAGGGGCGAGAAGAGTTGC
TGTGGATGCCAGGTGTGGGTTCATCAACTATAGACAATAAGAGGAGAGAAAGTCTTCTGG
GTTGGGGACATGGTAAGAGGGTGAGCAGTAGCTGGGCTGCCGGGAATAAAAGTCACACGT
AAAAGGGGCTCTTGTGTCTAGACTCCCAATATCAGATTTGATCACTAACCAGAATTTTTC

6594 TAACTTTGGGGCTGTGGTTAAGATGTATATATTTTGTGTATGACCTGCAGGTAAAACCAC

FIG.3H

```
          GGTGACCCAGTCAGTGGCAGATTCACTTAAGGCTGTCCTCTTAAAGTCACCACCCTCTTG
          CATTGGCCAGTGGAGGAAGATCTTTGATGATGAACCAACTATCATTAGAAGAGCTTTTTA
          CTCTTTGGGCAATTATATTGTGGCCTCCGAAATAGCTAAAGAATCTGCCAAATCTCCTGT
          GATTGTAGACAGGTAGGTATAAAGATGCCTTGAATTAGGCATTTTCTCCCTAATATATAA
          [-,G]
          TGTGTGTGTGTGTGTGTGTGTGTGTGTATACGTACATGTATATGCCAGGAAAAAAATT
          GTGTTTAAGTCAAACTGTTATTATGATAATAATAGGAATTCTCCTTATGAATTGTTAATT
          ACCTATACCAGGCATGGCATTTGCTAGAGAATTACATATATAATACTAGTATCTGGAACT
          ATAACTTGGGTAGGTGAATGTTACATGTTATTCCCAGTTTACTGATGAGAACTATAGATC
          TCAGAAAGGTAAAATAACTTGCCAAGGTAAGCTGGAAATAGCATACCGGGGACATTAATG

10987     AGCAGTTTGAAAAGATGTAGCAAAGTGGAAGATTGCTCCTGCCTACTGAAGAAACACTCA
          CAAAAGTTTCTTTCAGCTTCCTTCATACTGTTGCAAAATTTGAAAAACCTAAAGCCCATC
          AACAGGAGAAAGCATAAATAAATACCAGCATATGTATAAGGTGGAAAACCATACAGCAGC
          AGTGTAAATATGTTGTGCAATGCATGAGGGCACACACATTACAAAACCGTAATGTTGAAT
          GAAAATGTCAAGTTGCCAAAAGATGTTTATTCAGTAGAATACATGCAGTTATCCATACTG
          [C,T]
          AAAGAGTAGAGAGATGAAAGGATGTTGAACAGCAGATTTAAGAAAGTTGTTCTCTGGGAA
          GTAGAGATAAATGTGGTTGATAAGGAGATCAGAAGGACTTTGATTGTGATTGGATTATTT
          CTTAACATTTCTTGAATGTTCATTTTCTCCAAAGATTTACTTATTTTTGTAAACATAAGC
          AATTGCATTTAAGTATGGCATAGTTGGTAAAGAACATTAATCATAAATGTAAATCTCAAG
          GAATTTCAAGAAGTGAACACAGTCAAGTCACCTGTACCCAGATCAAGAATGAGAGCCTGC
```

Chromosome Map: chromosome 2

FIG.3I